Weighted Model Counting with Conditional Weights for Bayesian Networks (Supplementary Material)

Paulius Dilkas[1] **Vaishak Belle**[1]

[1]University of Edinburgh, Edinburgh, UK , p.dilkas@sms.ed.ac.uk, vaishak@ed.ac.uk

1 PROOFS

Theorem 1. *The function μ_ν is a measure.*

Proof. Note that $\mu_\nu(\bot) = 0$ since there are no atoms below $\bot$. Let $a, b \in 2^{2^U}$ be such that $a \wedge b = \bot$. By elementary properties of Boolean algebras, all atoms below $a \vee b$ are either below a or below b. Moreover, none of them can be below both a and b because then they would have to be below $a \wedge b = \bot$. Thus

$$\begin{aligned}\mu_\nu(a \vee b) &= \sum_{\{u\} \le a \vee b} \nu(u) = \sum_{\{u\} \le a} \nu(u) + \sum_{\{u\} \le b} \nu(u) \\ &= \mu_\nu(a) + \mu_\nu(b)\end{aligned}$$

as required. □

Theorem 3. *For any set U and measure $\mu\colon 2^{2^U} \to \mathbb{R}_{\ge 0}$, there exists a set $V \supseteq U$, a factorable measure $\mu'\colon 2^{2^V} \to \mathbb{R}_{\ge 0}$, and a formula $f \in 2^{2^V}$ such that $\mu(x) = \mu'(x \wedge f)$ for all formulas $x \in 2^{2^U}$.*

Proof. Let $V = U \cup \{f_m \mid m \in 2^U\}$, and $f = \bigwedge_{m \in 2^U} \{m\} \leftrightarrow f_m$. We define weight function $\nu\colon 2^V \to \mathbb{R}_{\ge 0}$ as $\nu = \prod_{v \in V} \nu_v$, where $\nu_v(\{v\}) = \mu(\{m\})$ if $v = f_m$ for some $m \in 2^U$ and $\nu_v(x) = 1$ for all other $v \in V$ and $x \in 2^{\{v\}}$. Let $\mu'\colon 2^{2^V} \to \mathbb{R}_{\ge 0}$ be the measure induced by ν. It is enough to show that μ and $x \mapsto \mu'(x \wedge f)$ agree on the atoms in 2^{2^U}. For any $\{a\} \in 2^{2^U}$,

$$\begin{aligned}\mu'(\{a\} \wedge f) &= \sum_{\{x\} \le \{a\} \wedge f} \nu(x) = \nu(a \cup \{f_a\}) \\ &= \nu_{f_a}(\{f_a\}) = \mu(\{a\})\end{aligned}$$

as required. □

Lemma 1. *Let $X \in \mathcal{V}$ be a random variable with parents $\mathrm{pa}(X) = \{Y_1, \dots, Y_n\}$. Then $\mathrm{CPT}_X\colon 2^{\mathcal{E}^*(X)} \to \mathbb{R}_{\ge 0}$ is such that for any $x \in \operatorname{im} X$ and $(y_1, \dots, y_n) \in \prod_{i=1}^n \operatorname{im} Y_i$,*

$$\mathrm{CPT}_X(T) = \Pr(X = x \mid Y_1 = y_1, \dots, Y_n = y_n),$$

where $T = \{\lambda_{X=x}\} \cup \{\lambda_{Y_i = y_i} \mid i = 1, \dots, n\}$.

Proof. If X is binary, then CPT_X is a sum of $2\prod_{i=1}^n |\operatorname{im} Y_i|$ terms, one for each possible assignment of values to variables $X, Y_1, \dots, Y_n$. Exactly one of these terms is nonzero when applied to T, and it is equal to $\Pr(X = x \mid Y_1 = y_1, \dots, Y_n = y_n)$ by definition.

If X is not binary, then $\left(\sum_{i=1}^m [\lambda_{X=x_i}]\right)(T) = 1$, and $\left(\prod_{i=1}^m \prod_{j=i+1}^m (\overline{[\lambda_{X=x_i}]} + \overline{[\lambda_{X=x_j}]})\right)(T) = 1$, so $\mathrm{CPT}_X(T) = \Pr(X = x \mid Y_1 = y_1, \dots, Y_n = y_n)$ by a similar argument as before. □

Lemma 2. *Let $\mathcal{V} = \{X_1, \dots, X_n\}$. Then*

$$\phi(T) = \begin{cases} \Pr(x_1, \dots, x_n) & \text{if } T = \{\lambda_{X_i = x_i}\}_{i=1}^n \text{ for some } (x_i)_{i=1}^n \in \prod_{i=1}^n \operatorname{im} X_i \\ 0 & \text{otherwise,} \end{cases}$$

for all $T \in 2^U$.

Proof. If $T = \{\lambda_{X = v_X} \mid X \in \mathcal{V}\}$ for some $(v_X)_{X \in \mathcal{V}} \in \prod_{X \in \mathcal{V}} \operatorname{im} X$, then

$$\begin{aligned}\phi(T) &= \prod_{X \in \mathcal{V}} \Pr\left(X = v_X \,\middle|\, \bigwedge_{Y \in \mathrm{pa}(X)} Y = v_Y\right) \\ &= \Pr\left(\bigwedge_{X \in \mathcal{V}} X = v_X\right)\end{aligned}$$

by Lemma 1 and the definition of a Bayesian network. Otherwise there must be some non-binary random variable $X \in \mathcal{V}$ such that $|\mathcal{E}(X) \cap T| \ne 1$. If $\mathcal{E}(X) \cap T = \emptyset$, then $\left(\sum_{i=1}^m [\lambda_{X=x_i}]\right)(T) = 0$, and so $\mathrm{CPT}_X(T) = 0$, and

Accepted for the 37th Conference on Uncertainty in Artificial Intelligence (UAI 2021).

$\phi(T) = 0$. If $|\mathcal{E}(X) \cap T| > 1$, then we must have two different values $x_1, x_2 \in \operatorname{im} X$ such that $\{\lambda_{X=x_1}, \lambda_{X=x_2}\} \subseteq T$ which means that $(\overline{[\lambda_{X=x_1}]} + \overline{[\lambda_{X=x_2}]})(T) = 0$, and so, again, $\mathrm{CPT}_X(T) = 0$, and $\phi(T) = 0$. □

Theorem 4. *For any $X \in \mathcal{V}$ and $x \in \operatorname{im} X$,*

$$(\exists_U (\phi \cdot [\lambda_{X=x}]))(\emptyset) = \Pr(X = x).$$

Proof. Let $\mathcal{V} = \{X, Y_1, \dots, Y_n\}$. Then

$$\begin{aligned}
(\exists_U (\phi \cdot [\lambda_{X=x}]))(\emptyset) &= \sum_{T \in 2^U} (\phi \cdot [\lambda_{X=x}])(T) \\
&= \sum_{\lambda_{X=x} \in T \in 2^U} \phi(T) \\
&= \sum_{\lambda_{X=x} \in T \in 2^U} \left(\prod_{Y \in \mathcal{V}} \mathrm{CPT}_Y \right)(T) \\
&= \sum_{(y_i)_{i=1}^n \in \prod_{i=1}^n \operatorname{im} Y_i} \Pr(x, y_1, \dots, y_n) \\
&= \Pr(X = x)
\end{aligned}$$

by:

- the proof of Theorem 1 by Dudek et al. [2020];
- if $\lambda_{X=x} \notin T \in 2^U$, then $(\phi \cdot [\lambda_{X=x}])(T) = \phi(T) \cdot [\lambda_{X=x}](T \cap \{\lambda_{X=x}\}) = \phi(T) \cdot 0 = 0$;
- Lemma 2;
- marginalisation of a probability distribution.

□

References

Jeffrey M. Dudek, Vu Phan, and Moshe Y. Vardi. ADDMC: weighted model counting with algebraic decision diagrams. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pages 1468–1476. AAAI Press, 2020. ISBN 978-1-57735-823-7. URL `https://aaai.org/ojs/index.php/AAAI/article/view/5505`.